采用 LDA 主题模型的国内知识流研究结构探讨：以学科分类主题抽取为视角*

王曰芬　傅　柱　陈必坤
(南京理工大学经济管理学院　南京 210094)

摘要：【目的】利用 LDA 主题模型探索分类视角下的主题提取与分布状态，全面深入地揭示学科知识结构和热点。【方法】以国内知识流领域为研究对象，选取 CNKI 和万方数据库中知识流相关文献为数据源，利用中图分类号将知识流研究的文献分成 11 个学科，借助 LDA 主题模型分别对这 11 个学科所包含的文献进行主题抽取，挖掘出不同学科中所包含的 20 个热点主题。【结果】分析 20 个热点主题，得到 11 个不同学科的热点主题内容及其所揭示的知识点。【局限】该方法没有同其他学科主题挖掘方法进行对比，分析得到的国内知识流研究热点也没有同已有相关文献中分析出的该领域热点进行对照。【结论】该方法能够更全面和深入地挖掘学科知识结构和研究热点。

关键词：知识流　LDA　主题抽取　学科分类　知识结构　研究热点
分类号：G254

1　引　言

当今科学研究的交叉渗透使大多学科领域在研究过程中不断呈现出多元交错的知识结构。尤其是新兴领域的研究，在多学科参与下其研究内容与结构更是复杂多样，为判别与掌握学科研究的内在构成带来一定的困扰。而目前有关学科知识的主题挖掘、结构探讨等研究，大部分是从全局的角度对领域进行分析和挖掘，无法全面、细致地揭示研究领域的知识结构。本文认为关于同样的研究领域，在跨学科参与中不同学科对同一领域的研究存在一定差异。

同时，随着科学技术的发展，传统的静态知识表示已然无法满足当前日新月异的知识活动需求，严重阻碍了知识管理的进一步实施和扩展，如何提出一种新的动态知识表示方法是当前知识管理学科亟待解决的问题之一。知识流中的知识是动态的，知识流不仅仅代表知识的运动，也能够解释知识如何在组织中移动[1]。知识流研究作为一个新兴的研究领域，正在吸引着不同学科领域的学者和管理人员的关注、研究和实践。自知识流提出以来，国内外学术界从不同角度对知识流进行了探讨，知识流领域的学术论文也一直保持着稳定的增长。因此，本文试图以知识流领域的研究为对象，借助 LDA 主题模型方法，深入挖掘与揭示不同学科下的知识流研究结构，以期进一步把握不同学科间知识流领域的研究现状、研究重点和发展方向，并为相关研究方法的应用提供参考借鉴。

2　相关研究现状

从科学文献中挖掘学科研究主题、学科结构和发展趋势，一般有两类方法：基于关键词的词频分析方法和共词分析方法[2-5]；基于 LDA 模型[6]、HDP 模型[7]等主题模型的主题挖掘方法。共词分析可以有效解决传统的词频分析方法的不足，通过统计关键词共现频次，构建共词矩阵进行聚类，以战略坐标图或主题网

通讯作者：王曰芬，ORCID：0000-0002-7143-7766，E-mail：yuefen163@163.com。

*本文系国家自然科学基金项目“新研究领域科学文献传播网络生长及对传播效果影响研究”(项目编号：71373124)和国家社会科学基金项目“学科结构与演化的可视化分析理论框架及应用研究”(项目编号：15CTQ035)的研究成果之一。

络的形式进行可视化。但是，共词分析也存在一些问题，如：关键词的选取具有主观性，如何获取表征论文主题的关键词常常成为共词分析的关键；由于文献主题词积累所引发的分析时滞性，导致不能够完全反映某一学科主题的发展趋势和最新生长点；难以反映主题词对间更深层次的语义关系。

基于对现有研究的总结与分析，本文重点对 LDA 主题模型的相关研究现状进行论述。中文文献中 LDA 主题模型的相关研究侧重基于 LDA 的文本主题抽取和演化、文本分类、文本分割等研究，如：张晨逸等[8]提出一种基于 LDA 的微博生成模型 MB-LDA，综合考虑了微博的联系人关联关系和文本关联关系，辅助进行微博的主题挖掘；关鹏等[9]将 LDA 主题模型与科学文献生命周期理论相结合，提出一种挖掘学科领域生命周期语义信息的方法；李文波等[9]提出一种附加类别标签的 LDA 模型(Labeled-LDA)，基于该模型可以在各类别上协同计算隐含主题的分配量，从而克服了传统 LDA 模型用于分类时强制分配隐含主题的缺陷，基于 Labeled-LDA 模型的文本分类新算法可以有效改进文本分类的性能；石晶等[11]基于 LDA 模型的进行文本分割研究；刁宇峰等[12]借鉴处理垃圾邮件的方法，针对 Blog 本身的特点，使用规则初步过滤垃圾评论，对剩余评论利用LDA这种能够提取文本隐含主题的产生式模型，对博客中的博文进行主题提取，并结合主题信息进行判断，从而识别 Blog 空间的垃圾评论。

英文文献中 LDA 主题模型的研究侧重于主题演化、语义识别、算法性能对比、算法改进等模型的应用和改进。如 Alsumait 等[13]提出在线主题模型，该主题可以自动获取主题模式和识别文本语料的新兴主题和时间变化；Lu 等[14]实现了系统对比分析 PLSA 和 LDA 两种概率主题模型在文档聚类、文本分类和专门检索三类文本挖掘任务中的性能；Zhai 等[15]提出约束性 LDA 方法用于产品评论观点挖掘中的产品特征组合实验，结果表明约束性 LDA 模型略优于原始 LDA 和最新的 mLSA 方法；Wei 等[16]提出基于 LDA 的文档模型，并在几个 TREC 语料集中测试，实验结果表明在信息检索中使用基于聚类的模型可以获得可靠的效率；Cao 等[17]研究 LDA 中最优主题结构和主题间距离的内在关联，提出基于密度的适应性最优 LDA 模型选择方法，该方法可以不需要调整主题个数而获得性能最佳的 LDA 主题模型。

综合以上分析，LDA 主题模型的相关应用研究主要涉及主题抽取和演化、文本分类、语义识别、算法改进等方面。其中，基于 LDA 模型的主题挖掘方法一般是利用科学文献全文、摘要等文本构建语料库，通过一系列主题模型抽取科学文献隐藏的主题结构，利用概率方法生成与该主题相关的一系列词项刻画主题内容。相比共词分析方法，该方法可以克服关键词选取的主观性问题，其结构组成是由一系列词语组成，而不是通过单一的共现词对聚类来刻画，通过这种描述方式能够深度挖掘主题语义信息，并通过量化的方式度量主题的强度以及主题之间的关系。

然而，目前基于 LDA 模型的主题挖掘方法基本上是直接利用 LDA 经典模型或者改进模型对全局语料进行主题抽取，该方式得到的主题能够描述领域中部分的热点主题，未能进一步从广度和深度上较为全面、深入地了解领域中各学科的热点主题。随着学科的交叉渗透，学科主题结构越来越复杂，学科主题的揭示也愈发困难，如何更为清晰有效地揭示学科主题是当前急需解决的问题之一。基于此，本文试图借助 LDA 主题模型，以学科分类为视角进行国内知识流领域的主题挖掘研究。

3 知识流领域的分学科主题抽取

本研究目的是通过 LDA 主题模型，以不同学科类别的主题抽取为视角揭示知识流研究领域的知识结构和主题热点内容。为此，通过文献数据库获取知识流领域相关文献的题录信息，对获取到的文献信息进行合并、去重等数据清洗操作，根据文献的中图分类号对文献信息进行学科分类，得到主题抽取的实验数据；对文献数据进行分词、模型训练等归一化处理，并设置 LDA 模型的相关参数；开展知识流领域的分学科主题抽取实验，筛选并确定知识流领域不同学科的热点主题。

3.1 文献获取和预处理

选取国内知识流领域为实验对象，实验数据检索自 CNKI 中的《中国学术期刊网络出版总库》(CAJD)和万方数据知识服务平台《中国学术期刊数据库》(CSPD)。检索策略是限定关键词或标题中精确包含“知识流”(时间截至 2014 年)。检索结果得到：来自 CNKI 数据库的文献有 487 篇，来自万方数据库的文献

有651篇。为获取不同学科下期刊文献，通过自编Java程序从两个数据中抓取题录信息，字段包括出版年份、标题、关键词、摘要、中图分类号。

为获取科学、规范的文献来源，并排除不相关文献的干扰，对获取到的文献进行如下处理：合并两个数据库中的期刊文献，在此基础上去重；删除关键词或标题字段中不包含“知识流”的期刊文献；删除标题、关键词、中图分类号不完整的文献，最终得到651条期刊文献题录。

在获得期刊文献题录后，根据中图分类号对文献进行学科分类。目前期刊文献中的分类号标注存在分类号标注层级不一致、一篇文献多个分类号等问题。针对该问题，在考虑文献的领域代表性、学科的辨识性和平衡性的基础上，并基于本次研究不考虑学科间研究交叉的问题，选取文献标注的首个分类号作为文献的主要类目，最终选择二级目录作为文献分类的标准。对所获得文献的分类号进行统计，其分布情况如表1所示：

表1　文献的分类号分布情况

序号	分类号	数量	序号	分类号	数量
1	F2	249	25	F8	2
2	G3	66	26	N0	2
3	F0	66	27	F3	2
4	G2	37	28	I3	1
5	F4	33	29	C26	1
6	G6	28	30	I2	1
7	TP3	24	31	C03	1
8	C93	24	32	O1	1
9	F1	18	33	C81	1
10	TP1	15	34	R4	1
11	G4	13	35	E2	1
12	F7	13	36	R5	1
13	G7	10	37	R2	1
14	TB4	5	38	G1	1
15	R1	4	39	TU9	1
16	N9	3	40	TH1	1
17	F5	3	41	B9	1
18	O4	2	42	D4	1
19	G8	2	43	V1	1
20	C96	2	44	F9	1
21	D9	2	45	F6	1
22	D6	2	46	U6	1
23	B8	2	47	K9	1
24	C91	2			

(注：C93因其二级目录无对应的分类名称，故以其三级类目作为分类基准。)

从表1中看出，前13个类的文献数量为596，占全部文献数据(651)的 91.55%，能够代表知识流领域，且除了 F2 类以外，其他分类目录的文献分布较为平衡。因此，本文选择前13个类目的596篇文献作为本实验的数据。因G4类包含G6类和G7类，本文将这三个类统归为G4，得到11个学科分类如表2所示：

表2　Top11 分类目录名称

类别序号	分类号	数量	学科分类名称
1	F2	249	经济计划与管理
2	G3	66	科学、科学研究
3	F0	66	经济学
4	G4{G6}{G7}	51	教育
5	G2	37	信息与知识传播
6	F4	33	工业经济
7	TP3	24	计算技术、计算机技术
8	C93	24	管理学
9	F1	18	世界各国经济概况、经济史、经济地理
10	TP1	15	自动化基础理论
11	F7	13	贸易经济

3.2　文献数据归一化处理和 LDA 参数设置

利用中国科学院计算技术研究所分词工具包ICTCLAS[18]实现中文分词、去除停用词等自然语言处理规范化过程，获得实验用语料库，基于开源包JGibbLDA[19]实现 LDA 主题模型的参数训练。

对于各类别数据，在进行LDA主题模型参数设置时，使用 Gibbs Sampling 进行参数后验估计，设置迭代次数为1 000次，超参数设置为α=0.01、β=0.05，各类别数据的主题数目 $T_{i\,(i=1,2,3,\cdots,11)}$暂且全部设置为20个。

LDA主题抽取后获得两个重要文档，一个是文档–主题分布文档命名为<model_name>.theta，可用来计算主题强度；一个是主题–词分布文档命名为<model_name>.phi，用来表示主题及其内容。

3.3　实验结果及分析

实验结果得到11个学科的文档–主题分布和主题–词分布，如图1和表3所示。在各学科分类中选取主题强度大于0.1的主题作为热点，其中C93类和G3类的主题强度均小于 0.1，为了学科主题分布的平衡性，本文将C93类的主题10以及G3类的主题7和主题10纳入实验结果以便对比分析，按照各分类的文献数量排序，不同类别的主题数量分布如表4所示。

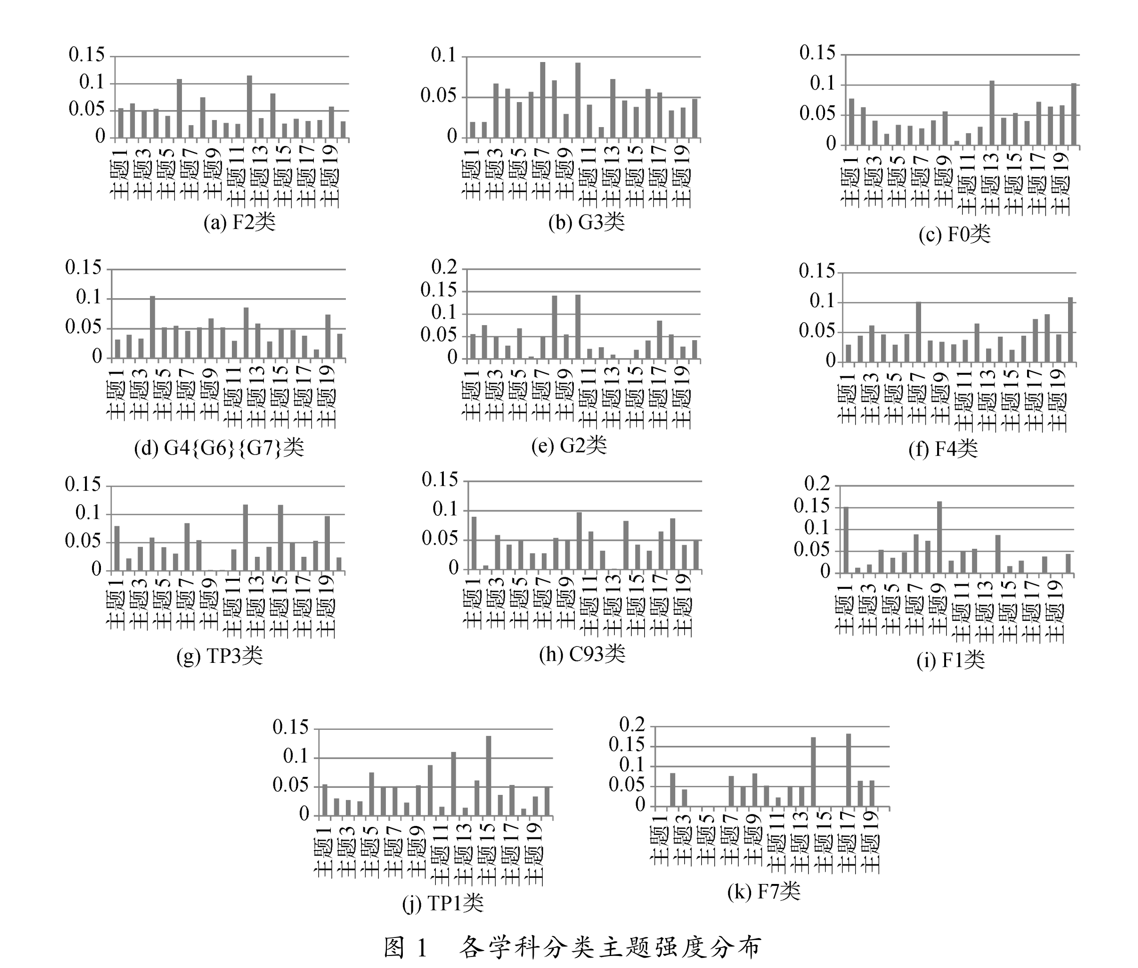


图 1　各学科分类主题强度分布

表 3　各学科分类的主题展示(省略词项的概率)

学科分类	主题标识	与研究主题最相关的词项(前 20 个)
F2	主题 12：业务流程知识流管理	知识流、知识管理、业务流程、组织知识、知识流程、知识共享、知识经济时代、知识流分析、知识流管理、知识流模型、知识创新、内部知识、知识活动、信息技术、组织结构、知识资源、企业业务流程、知识管理系统、业务流程重组、知识地图
	主题 6：跨国公司知识流动	知识流动、跨国公司、竞争优势、知识创造、技术创新、核心竞争力、网络组织、影响因素、知识共享、知识存量、知识管理、知识网络、知识创新、知识流量、知识资源、知识基础、核心企业、学习能力、企业竞争、知识整合
G3	主题 7：知识流测度	知识流动、影响因素、知识流动过程、引文分析、知识链、知识流动模型、知识链组织、社会网络、知识进化、知识转化、知识流动测度、专利引证、延展性、知识流动效率、动态确定、流动效率、重要性、引文分析学、知识活动、引文机制
	主题 10：技术创新知识流动	知识流、知识流动、技术创新、技术创新能力、知识存量、知识理论、知识经济时代、知识转化、作用机制、知识管理、知识获取、技术能力、动态化、关系、社会互动、环境压力、组织知识、技术知识流动、技术转移、产学研
F0	主题 13：产业集群知识流动	知识流动、产业集群、竞争优势、知识网络、区域经济、知识链、价值链、知识经济、知识守门者、超本地知识流动、产业集群升级、开放性、关键知识、知识扩散、县域特色、创新型产业集群、均衡模型、技术创新扩散、测量模型、运筹学
	主题 20：区域创新知识流动	知识流动、区域创新系统、竞争力、创新能力、重要性、知识流动模型、机理研究、分析知识、空间集聚、知识、创新链、区域经济、F1 产业、复杂产品系统、依赖性、模型、科技服务业、国民经济、子系统、知识经济时代

(续表)

学科分类	主题标识	与研究主题最相关的词项(前 20 个)
G4 {G6} {G7}	主题 4: 高职院校知识创新	知识流、核心竞争力、知识管理、高职院校、知识创新、知识库、激励机制、知识管理体系、知识链、价值增值、知识共享、知识流动、教学活动、组织结构、知识产权、高校核心竞争力、路径选择、知识网格、战略、路径
G2	主题 10: 图书馆中显隐性知识转化	知识流、知识管理、图书馆、知识流动、隐性知识、显性知识、知识、SECI、Living Library、知识共享、SECI 模型、知识增值、学科馆员、知识经济时代、图书馆讲座、信息流、人才流、思想流、终身学校、基层图书馆
	主题 8: 图书馆和企业知识传播管理	知识管理、知识流程、图书馆、业务流程、知识链、企业管理、构成要素、知识共享、管理模式、知识创新、企业知识管理模型、竞争力、知识链模型、知识链管理、竞争能力、电子政务、政务知识流程、SECI、知识流、Living Library
F4	主题 20: 知识流动模型研究	知识流动、知识共享、知识溢出、知识、知识传递、知识流动模型、价值链、流动方式、分析研究、知识转移、战略联盟、知识流动循环、知识势差、知识流动过程、知识创新、核心能力、企业知识流动、企业、获取、模型
	主题 7: 产业集群和知识创新	知识流动、竞争优势、产业集群、网络结构、观念转变、西部、工业、创新要素、模糊数学、评价指标体系、针对性、集群创新、突破性、模糊综合评价、高新技术产业集群、分析框架、企业行为、共同体、知识存量、政治学
TP3	主题 12: 工程设计知识流控制建模	知识流、工作流、表示方法、集成建模、知识流引擎、协同设计、控制技术、资源空间模型、集成控制、逻辑关系、检索、推送、过程控制、工作流过程元模型、集成建模、知识集成、学生管理、耦合建模、一体化设计、可靠性
	主题 15: 基于 Petri 网的知识流建模	知识流、知识流模型、Petri 网、着色 Petri 网、显性知识、知识生命周期、知识管理、隐性知识、生命周期、隐性知识流、仿真分析、模型仿真、知识流管理、协作图、供应链、业务流程、集成系统、耦合建模、一体化设计、可靠性
C93	主题 10: 组织学习知识交流	知识流动、组织学习、知识流、知识流模型、知识创造、知识交流、知识创新、模型、知识共享、双元平衡、组织学习情境、探索式学习、利用式学习、计算机仿真、影响因素、管理咨询服务、知识转移模式、咨询业、知识转移、业务流程
F1	主题 9: 国家/区域创新系统	知识流动、区域创新系统、创新系统、区域经济、知识、动力机制、经济绩效、相关性、产业结构、经济增长、国家创新系统、创新主体、经济运行、创新绩效、创新、稳定度、创新能力、结果显示、技术创新能力、密切相关
	主题 1: 国家/区域创新体系	知识流动、区域创新体系、影响因素、区域创新、技术创新、服务机构、包容性创新、长三角、对策、国家创新体系、创新体系理论、区域创新系统、稳定度、创新能力、结果显示、技术创新能力、密切相关、知识流、演化、产业结构
TP1	主题 15: 业务协作知识流	知识管理、知识流、工作流、认知协作、知识管理研究、角色协作、组织结构、知识流动效率、知识共享、表示层、知识本体、知识流分析、知识管理系统、流程协调、异常处理、参与者、Agent、半主动、信牌、政府信息发布
	主题 12: 自动化控制知识流建模	知识管理、知识流、Petri 网模型、知识流动、知识流模型、知识流动效率、复杂性、Petri 网、流程控制、管理理论、分布情况、可能性、知识分布、知识存量、管理人员、表示层、知识本体、知识流分析、知识管理系统、流程协调
F7	主题 17: 服务创新	知识流、信息流、资金流、服务创新、知识流动、服务业、政策性建议、电子商务、运行平台、互联网、运作模式、物流、契约、实物形态、商品流、捷安、企业转型、内外部、公共服务、外部环境
	主题 14: 服务外包	服务外包、知识流程外包、信息技术外包、科技服务业、产生背景、联动机制、业务流程外包、企业转型、内外部、公共服务、外部环境、因子分析、产业链、知识流、政策扶持、服务创新、知识流动、服务业、政策性建议、Blotto 博弈

表 4 Top11 学科分类选取的主题个数

序号	分类号	数量	主题个数
1	F2	249	2
2	G3	66	2
3	F0	66	2
4	G4{G6}{G7}	51	1
5	G2	37	2
6	F4	33	2
7	TP3	24	2
8	C93	24	1
9	F1	18	2
10	TP1	15	2
11	F7	13	2
	总计	596	20

4 结果与讨论

结合相关文献按照学科对国内知识流的研究结构及其热点进行逐一解读和阐释。

(1) F2: 经济计划与管理

①业务流程知识流管理。该主题是基于业务流程的组织内部知识流管理研究，主要是利用信息技术，基于业务流程开展知识共享、知识创新、知识流管理、知识流建模等方向的研究。如蔡强等[20]基于业务流程和其内部知识特征，提出一种从知识流优化角度进行流程重组的方法。赵涛等[21]从隐性知识共享的角度出发，揭示隐性知识的内涵，通过研究企业业务流程中的关键驻点，建立隐性知识流的动态扩展模型，分析企业业务流程之间隐性知识流转移。

②跨国公司知识流动。该主题是跨国公司研发(R&D)活动、国际化合作等方面的知识流动研究。楚天骄等[22]认为跨国公司研发机构与本土互动的效果是由跨国公司研发机构的性质、地方创新主体的技术水平及其与跨国公司研发机构之间的技术落差、地方研发人员的素质和丰富程度、知识产权保护力度、地理接近程度等因素综合作用的结果。元利兴等[23]对跨国公司全球 R&D 活动中的知识流动机制进行研究，主要包括跨国公司的知识网络结构、跨国公司全球 R&D 活动中的知识流动模型及其影响因素。

(2) G3: 科学、科学研究

①知识流测度。该主题是知识链中知识流的流动原因、流动效率影响因素等方面的研究。并且，在知识链的知识流动测度研究中，需要结合社会网络、引文分析、专利引证等研究方法。知识链中知识流动研究有：如赵力焓等[24]分析知识链组织之间知识流动的原因、本质、方式、特征等因素，研究知识链组织之间的知识流动过程，构建包含五步骤的知识链组织之间知识流动模型，提出促进知识链组织之间知识流动的建议。吴绍波等[25]认为知识链组织的合作过程是一个组织之间知识流动的过程。以信任为基础的冲突协调机制能增加知识流动中的透明度，使知识链组织增加合作中的专用性投资以及保持沟通的开放性，从而能弥补契约机制对知识链组织冲突协调的不足。

结合社会网络、引文分析、专利引证等方法的知识流相关研究有：梁永霞等[26]认为可以把文献引用的过程抽象为知识流动的过程，文献引用的过程都是在前人知识的基础上知识进化的过程，是知识的选择、遗传和变异的过程，也是知识的生产、传播和应用的过程。安宁等[27]以专利引证作为研究知识流入的手段，构建地域、技术领域、研发主体和时滞 4 个维度的指标体系，通过实证分析总结出目前我国医药技术领域知识流入的特点。钟琦等[28]从企业技术网络和社会网络对知识流动的影响分析入手，探讨在两种网络共同作用下的知识流动的概念、特征，提出技术网络和社会网络下知识流动的不同模式，在此基础上构建企业内部知识流动体系模型并进行实例分析。

②技术创新知识流动。该主题是组织中技术创新过程的知识流研究，技术创新是组织发展的必要条件之一，组织中知识流动效率、知识转化能力、知识存量对提升技术创新能力具有关键作用。如张培富等[29]通过分析知识流的内涵与本质，把握知识流与群体社会互动的内在联系，提出我国技术创新工作中促进知识流动与技术创新群体社会互动的若干对策。高建新[30]认为组织的知识转化是组织自身发展的内在要求；而在适应环境压力的过程中，组织形成了自己的知识特征；组织发展的过程是组织知识的动态转化与组织的知识特征相互作用的过程。通过这个过程，组织的知识特征与组织形态相匹配。

(3) F0: 经济学

①产业集群知识流动。该主题是产品集群过程中的知识流动研究，在集群企业中需要借助知识网络和知识链实现技术创新的扩散。如宜建军等[31]介绍了产业集群的特征，对知识流进行概述，建立产业集群内外知识流模型，并就基于知识流的产业集群知识管理过程进行讨论。林岩等[32]系统地讨论并检验上游企业利用上行知识流促进自身创新的问题，并分析立即作用和延迟作用两种促进作用。选取美国汽车生产行业中的零部件供应商为研究样本，实验结果验证了汽车生产行业中价值链下游是上游非常重要的知识源。

②区域创新知识流动。该主题主要是区域创新系统中创新扩散、知识外溢、知识共享等现象的知识流动机理研究。如孟庆敏等[33]通过对区域创新系统结构的分析，建立区域创新系统的知识流动模型，并分析科技服务业的运行机理。顾新[34]借鉴国家创新系统理论和方法，从知识流动、产业集聚、空间集聚三个方面分析了区域创新系统的运行机理。

(4) G4{G6}{G7}：教育

高职院校知识创新。该主题是基于知识流的高职院校核心竞争力研究，从知识流的角度看知识链和学习链是高职院校的核心竞争力，知识流动和知识创新实现知识的价值增值。如余霞等[35]回顾了高职院校核心竞争力评价研究文献，从知识流的视角构建了包括31个三级指标、10个二级指标、2个一级指标在内的高职院校核心竞争力评价指标体系，并对各级指标进行分析。余霞[36]认为高职院校核心竞争力的提升要对其知识流进行重组和改造，也就是降低知识转移成本，提高知识转移和创新效率，从而提升其知识价值，促进知识增值。

(5) G2：信息与知识传播

①图书馆显隐性知识转化。该主题是图书馆知识管理，侧重基于SECI模型的知识流动和转化，为用户提供知识服务，实现知识的增值。如田丁等[37]在讨论知识库功能与作用的基础上，设计构建一种以知识库建设为基础的图书馆知识流通系统，为图书馆实现以用户知识需求为中心的知识组织与服务的知识管理提供有效的途径。蔡璇璇[38]基于SECI模型和知识流的Living Library活动，提出活动不仅实现了SECI模型中隐性知识和显性知识的转化，而且活动中知识流的作用产生了知识增值效用。

②图书馆和企业知识传播管理。该主题的内容主要是基于知识链的企业或图书馆业务流程和知识流程管理。如冯锐等[39]从外包服务驱动因素、外包活动要素两个方面，对高校图书馆业务流程外包与知识流程外包进行了对比分析，总结了两种外包模式的异同点。江亮[40]将知识管理思想融入到图书馆业务流程中，提出图书馆知识流程的概念，阐述图书馆知识流程的构成要素以及支撑该流程循环的条件。

(6) F4：工业经济

①知识流动模型研究。该主题是知识流动模型研究，包括知识共享、知识传递、流动方式、知识转移等内容。如王月平[41]构建了战略联盟的知识流动循环模型，认为知识势差导致战略联盟知识流动过程的形成，知识流动过程由知识共享阶段、知识创新阶段和知识吸收阶段构成，知识流动的最终结果是导致组织核心能力的提高和知识势差的弥补。赵顺龙等[42]从纵向合作中知识的流动方式出发，指出企业在纵向合作中对核心知识进行保护的两条路径：警惕核心知识被上下游合作伙伴窃取及预防核心知识被竞争对手获取。

②产业集群和知识创新。该主题是产业集群、网络组织中知识流动和知识创新研究。如阎海峰[43]从影响知识流动的5个主要因素入手，以跨国网络结构为例，探讨网络结构中知识流动的主要角色，进而从信息政治学的角度提出网络结构和有效激励机制是对付知识流动障碍、提高流动率的重要途径。丁立群等[44]运用模糊数学原理，构建知识流动促进产业集群创新的评价指标体系，剖析区域内外知识流动影响产业集群创新的重要因子，并针对性地提出扩散、集成、提升和转化产业集群显性和隐性知识。

(7) TP3：计算技术、计算机技术

①工程设计知识流控制建模。该主题是基于工作流的工程设计知识流动及其控制建模研究。如王生发等[45]将资源空间模型引入知识流中的知识描述，将知识流与工作流进行集成，提出知识流与工作流的集成控制面板，实现协同设计中的知识流控制。张晓刚等[46]提出一种知识流建模方法，通过5类知识流单元对知识传递与重用、人员协作与交流进行表示。针对知识流中的动态因素，研究基于资源约束、知识需求变化和时间约束的知识流控制方法，以实现自适应的知识流控制。

②基于Petri网的知识流建模。该主题是组织中基于Petri网的显隐性知识流动建模研究。如朱卫未等[47]采用着色Petri网以不同颜色的托肯表示不同类型的知识，构建基于组织知识生命周期的知识流模型。运用CPNTools软件对模型进行仿真分析，重点对Petri网中各节点的有界性和活性进行分析。丁漪杰等[48]认为基于Petri网的知识流建模，可以形象描述知识异步并发，而着色Petri网克服了普通Petri网的缺点，并且可以更加详细地描述企业知识流中的各种知识。

(8) C93：管理学

组织学习知识交流。该主题是组织中学习机制对组织知识流动的影响研究。如赵晨等[49]为明确领导个人学习对组织学习成效影响的内在机制，提出领导可以根据组织学习情境特征，采取与之相适应的探索式学习或利用式学习，调整组织内部的纵向知识流动和横向知识流动，从而在领导个人层面实现两种学习方式的双元平衡，在组织整体层面提高组织学习成效。张成考等[50]对虚拟企业中的知识流动规律和组织间学习机制进行深入研究，并对丰田知识联盟中影响知识流动与组织间学习的4个因素进行实证分析。

(9) F1：世界各国经济概况、经济史、经济地理

①国家/区域创新系统。该主题是国家或区域创新系统的知识流动特征及其对产品结构、经济增长的影响研究。如李正风等[51]讨论了关于创新系统研究中运用知识流分析方法的两种视角，以及不同视角下的不同切入方式。张瑞军等[52]认为创新系统中的知识流动是创新主体间相互作用的基本方式，作者借鉴国家创新系统研究的思路，从区域创新系统的角度出发，对知识流动与经济增长的相关性进行实证分析。

②国家/区域创新体系。该主题是国家或区域的技术创新体系知识流动研究。如魏奇锋等[53]通过分析长三角区域创新体系知识流动的特点，建立分层螺旋的知识流动结构模型，从环境、主体与客体等方面讨论了长三角区域创新体系中知识流动的影响因素。黄西川等[54]以我国区域技术创新调查为案例，通过实证分析得出区域创新主体间的知识流动水平与其技术创新绩效之间存在较强的正相关。

(10) TP1：自动化基础理论

①业务协作知识流。该主题是业务协作过程中认知协作和知识流程协调研究。如于腾等[55]以角色为认知主体进行抽象表示，讨论知识流中的角色协作，依据知识流的流向，将角色协作区分为相邻任务角色协作和相同任务角色协作。窦万春等[56]为了有效地对认知科学中基于群学习方式的认知协作进行研究，探讨基于本体描述的认知基础环境。在此基础上，对知识流关系进行形式化定义，并利用 Markov 决策过程理论分析研究了知识应用集成环境下协同认知的过程逻辑。

②自动化控制知识流建模。该主题是基于 Petri 网的知识流动建模和知识流程控制研究。周密等[57]把 Petri 网理论用在对知识流进行刻画中，提出知识流的 Petri 网模型，能够描述知识分布的拓扑结构、知识存量、一定时期的知识流量和流速，使得管理人员便于监控和管理知识流动，更好地实施知识管理。周密等[58]利用 Petri 网作为工具对知识的流动进行建模，从而得到知识流的概念模型。

(11) F7：贸易经济

①服务创新。该主题是知识流对服务业的升级和创新影响因素研究。如刘小军等[59]针对知识密集型服务业，分析服务创新体系中知识流动的 4 种基本途径，进一步阐述服务创新知识流所特有的一些障碍，并提出克服这些障碍的政策性建议。章依凌等[60]实证分析影响嘉兴外贸生产企业发展的内外部因素，并在此基础上提出通过政府公共服务和引导为转型企业营造最佳外部环境的对策建议。

②服务外包。该主题是服务业的知识流程外包、知识服务外包等相关研究。如王伟军等[61]分析知识流程外包中的知识产权授权主体，提出我国知识流程外包产业中知识产权的授权模式及各模式的选择标准，并对知识流程外包企业——易唯思中国商务咨询有限公司的知识产权授权模式进行具体剖析。朱胜勇[62]从产值结构、就业结构、离岸外包市场、离岸外包行业分布 4 个方面综述印度服务外包发展的状况，比较中印两国服务外包发展的差异，并为我国服务外包产业的发展提出建议。

5 结 语

本文以学科主题抽取为视角，利用LDA主题模型抽取了国内知识流研究领域中 11 个学科的 20 个热点主题，分析了不同学科热点主题的研究内容、应用领域和主要研究方法。实验结果显示国内知识流研究分布较广，文献数量最多的是 F2 类(经济计划与管理)，最少的是 F7 类(贸易经济)，学科主题抽取不仅可以全面、深入地揭示国内知识流研究领域的研究结构，并且学科主题抽取得到的 20 个主题基本是不交叉的，这也可以从侧面反映出跨学科研究领域在不同学科的研究热点存在差异性。因此，本文认为学科主题抽取对于领域研究热点发现、主题挖掘、结构探测具有重要的现实意义。当然，本文研究方法还存在不足之处，如主题抽取过程中主题个数的确定、热点主题的选取等问题，这都需要在今后的研究中加以改进。

参考文献：

[1] Nissen M E. Harnessing Knowledge Dynamics: Principled Organizational Knowing &Learning [M]. IRM Press, 2006.

[2] 王曰芬，宋爽，卢宁，等. 共现分析在文本知识挖掘中的应用研究[J]. 中国图书馆学报，2007，33(2)：59-64. (Wang Yuefen, Song Shuang, Lu Ning, et al. Application of Co-occurrence Analysis in Text Knowledge Mining [J]. Journal of Library Science in China, 2007, 33 (2): 59-64.)

[3] Hu C P, Hu J M, Deng S L, et al. A Co-word Analysis of Library and Information Science in China [J]. Scientometrics, 2013, 97(2): 369-382.

[4] Romo-Fernández L M, Guerrero-Bote V P, Moya-Anegón F. Co-word Based Thematic Analysis of Renewable Energy (1990-2010) [J]. Scientometrics, 2013, 97(3): 743-765.

[5] Guardo M C D, Harrigan K R. Mapping Research on Strategic Alliances and Innovation: A Co-citation Analysis [J]. Journal of Technology Transfer, 2012, 37(6): 789-811.

[6] Blei D M, Ng Y A, Tordan M. Latent Dirichlet Allocation [J]. Journal of Machine Learning Research, 2003, 3: 993-1022.

[7] 刘少鹏，印鉴，欧阳佳，等. 基于 MB-HDP 模型的微博主题挖掘[J]. 计算机学报，2015，38(7)：1408-1419. (Liu Shaopeng, Yin Jian, Ouyang Jia, et al. Topic Mining from Microblogs Based on MB-HDP Model[J]. Chinese Journal of Computers, 2015, 38(7): 1408-1419.)

[8] 张晨逸，孙建伶，丁轶群. 基于 MB-LDA 模型的微博主题挖掘[J]. 计算机研究与发展，2012，48(10)：1795-1802. (Zhang Chenyi, Sun Jianling, Ding Yiqun. Topic Mining for Microblog Based on MB-LDA Model [J]. Journal of Computer Research and Development, 2012, 48(10): 1795-1802.)

[9] 关鹏，王曰芬. 基于 LDA 主题模型和生命周期理论的科学文献主题挖掘[J]. 情报学报，2015，34(3)：286-299. (Guan Peng, Wang Yuefen. Topic Mining in Scientific Based on LDA Topic Model and Life Cycle Theory [J]. Journal of the China Society for Scientific and Technical Information, 2015, 34(3): 286-299.)

[10] 李文波，孙乐，张大鲲. 基于 Labeled-LDA 模型的文本分

类新算法[J]. 计算机学报, 2008, 31(4): 620-627. (Li Wenbo, Sun Le, Zhang Dakun. Text Classification Based on Labeled-LDA Model [J]. Chinese Journal of Computers, 2008, 31(4): 620-627.)

[11] 石晶, 胡明, 石鑫, 等. 基于 LDA 模型的文本分割[J]. 计算机学报, 2008, 31(10): 1865-1873. (Shi Jing, Hu Ming, Shi Xin, et al. Text Segmentation Based on LDA Model [J]. Chinese Journal of Computers, 2008, 31(10): 1865-1873.)

[12] 刁宇峰, 杨亮, 林鸿飞. 基于 LDA 模型的博客垃圾评论发现[J]. 中文信息学报, 2011, 25(1): 41-47. (DiaoYufeng, Yang Liang, Lin Hongfei. LDA-Based Opinion Spam Discovering [J]. Journal of Chinese Information Processing, 2011, 25(1): 41-47.)

[13] Alsumait L, Barbará D, Domeniconi C, et al. On-line LDA: Adaptive Topic Models for Mining Text Streams with Applications to Topic Detection and Tracking[C]. In: Proceedings of the 8th IEEE International Conference on Data Mining, 2008: 3-12.

[14] Lu Y, Mei Q, Zhai C X. Investigating Task Performance of Probabilistic Topic Models: An Empirical Study of PLSA and LDA [J]. Information Retrieval, 2011, 14(2): 178-203.

[15] Zhai Z, Liu B, Xu H, et al. Constrained LDA for Grouping Product Features in Opinion Mining [A]. // Advances in Knowledge Discovery & Data Mining [M]. Springer Berlin Heidelberg, 2011.

[16] Wei X, Croft W B. LDA-Based Document Models for ad-hoc Retrieval [C]. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 2006: 178-185.

[17] Cao J, Xia T, Li J, et al. A Density-based Method for Adaptive LDA Model Selection [J]. Neurocomputing, 2009, 72(7-9): 1775-1781.

[18] ICTCLAS [CP/OL]. [2015-12-11]. http://www.oschina.net/p/freeictclas.

[19] JGibbLDA-v.1.0 [CP/OL]. [2015-12-11]. http://sourceforge.net/projects/jgibblda/.

[20] 蔡强, 韩东梅, 李海生, 等. 基于知识流优化的业务流程重组[J]. 华中科技大学学报: 自然科学版, 2013, 41(S2): 33-37. (Cai Qiang, Han Dongmei, Li Haisheng, et al. Business Process Reengineering Based on Knowledge Flow Optimization [J]. Journal of Huazhong University of Science and Technology: Natural Science Edition, 2013, 41(S2): 33-37.)

[21] 赵涛, 曾金平. 企业隐性知识流动态扩展模型分析[J]. 科学学研究, 2005, 23(4): 536-539. (Zhao Tao, Zeng Jinping. Analysis on an Extended Dynamic Model of Tacit Knowledge-flow Through the Enterprise [J]. Studies in Science of Science, 2005, 23(4): 536-539.)

[22] 楚天骄, 杜德斌. 跨国公司研发机构与本土互动机制研究[J]. 中国软科学, 2006(2): 127-132. (Chu Tianjiao, Du Debin. The Interactional Mechanism of MNEs R&D Institution and Indigenous Principal Parts of Host RIS [J]. China Soft Science, 2006(2): 127-132.)

[23] 元利兴, 宣国良. 跨国公司全球 R&D 活动中的知识流动机制研究[J]. 科技进步与对策, 2003, 20(8): 34-35. (Yuan Lixing, Xuan Guoliang. Research on Mechanism of Knowledge Flow in Multinational Company's Global R&D Activities [J]. Science & Technology Progress and Policy, 2003, 20(8): 34-35.)

[24] 赵力焓, 石娟, 顾新. 知识链组织之间知识流动的过程研究[J]. 情报杂志, 2010, 29(7): 70-73. (Zhao Lihan, Shi Juan, Gu Xin. Study of Knowledge Flow Process Within Knowledge Chain[J]. Journal of Intelligence, 2010, 29(7): 70-73.)

[25] 吴绍波, 顾新, 彭双, 等. 知识链组织之间的冲突与信任协调: 基于知识流动视角[J]. 科技管理研究, 2009, 29(6): 325-327. (Wu Shaobo, Gu Xin, Peng Shuang, et al. The Knowledge Chains Conflicts and Trust Coordination: From the View of Knowledge Flow[J]. Science and Technology Management Research, 2009, 29(6): 325-327.)

[26] 梁永霞, 刘则渊, 杨中楷. 引文分析学的知识流动理论探析[J]. 科学学研究, 2010, 28(5): 668-674. (Liang Yongxia, Liu Zeyuan, Yang Zhongkai. Explore and Discussion on Knowledge Flowing Theory of Citation Analytics[J]. Studies in Science of Science, 2010, 28(5): 668-674.)

[27] 安宁, 刘娅. 基于专利引证的我国医药技术领域知识流入研究[J]. 科技管理研究, 2010, 30(18): 168-172. (An Ning, Liu Ya. Study on Knowledge Inflow in the Field of Drugs & Medical Technology in China Based on Patent Citations[J]. Science and Technology Management Research, 2010, 30(18): 168-172.)

[28] 钟琦, 汪克夷, 冯桂平. 网络视角下的企业内部知识流动体系研究[J]. 科技进步与对策, 2009, 26(24): 176-179. (Zhong Qi, Wang Keyi, Feng Guiping. Study on Inner Knowledge Flow System in Enterprises in a Network View[J]. Science & Technology Progress and Policy, 2009, 26(24): 176-179.)

[29] 张培富, 李艳红. 知识流与技术创新的群体社会互动[J]. 科技管理研究, 2004, 24(4): 105-109. (Zhang Peifu, Li Yanhong. Knowledge Flow and Social Groups' Interactions of

Technological Innovation [J]. Science and Technology Management Research, 2004, 24(4): 105-109.)

[30] 高建新. 组织的知识转化、知识特征及其作用机制[J]. 软科学, 2003, 17(3): 2-4. (Gao Jianxin. Knowledge Conversion, Knowledge Trait of Organization and Their Interaction Mechanism [J]. Soft Science, 2003, 17(3): 2-4.)

[31] 宜建军, 岳琳. 基于知识流的产业集群知识管理过程研究[J]. 情报杂志, 2010, 29(3): 107-109. (Yi Jianjun, Yue Lin. Study on the Industrial Colony Knowledge Management Process Based on the Knowledge Flow [J]. Journal of Intelligence, 2010, 29(3): 107-109.)

[32] 林岩, 陈燕, 李剑锋. 价值链中的上行知识流对供应商的促进作用——以汽车生产行业为例[J]. 科学学研究, 2010, 28(8): 1181-1191. (Lin Yan, Chen Yan, Li Jianfeng. Effects of the Upwards Knowledge Flow on Suppliers in the Supply Chain: An Example in the Auto Manufacturing Industry[J]. Studies in Science of Science, 2010, 28(8): 1181-1191.)

[33] 孟庆敏, 梅强. 科技服务业在区域创新系统中的功能定位与运行机理研究[J]. 科技管理研究, 2010, 30(8): 74-75. (Meng Qingmin, Mei Qiang. Study on Function Orientation and Mechanism of Science and Technology Service Industry in Regional Innovation System [J]. Science and Technology Management Research, 2010, 30(8): 74-75.)

[34] 顾新. 区域创新系统的运行[J]. 中国软科学, 2001(11): 105-108. (Gu Xin. The Operation of Regional Innovation System [J]. China Soft Science, 2001(11): 105-108.)

[35] 余霞, 张智光. 基于知识流的高职院校核心竞争力评价指标体系研究[J]. 教育与职业, 2012(12): 12-14. (Yu Xia, Zhang Zhiguang. Research on Index System of Core Competitiveness Evaluation of Higher Vocational Colleges Based on Knowledge Flows[J]. Education and Vocation, 2012(12): 12-14.)

[36] 余霞. 知识流与高职院校核心竞争力耦合研究[J]. 现代教育管理, 2010(10): 79-82. (Yu Xia. A Study on the Coupling of Knowledge Flow and the Core Competitiveness of Higher Vocational Colleges[J]. Modern Education Management, 2010(10): 79-82.)

[37] 田丁, 方曙. 基于知识库建设的图书馆知识流通系统的构建[J]. 图书情报工作, 2004, 48(1): 66-69. (Tian Ding, Fang Shu. A Study on the System of Library Knowledge Delivery Based on the Knowledge Database [J]. Library and Information Service, 2004, 48(1): 66-69.)

[38] 蔡璇璇. 基于 SECI 和知识流的 Living Library 活动[J]. 科技创新与生产力, 2013(8): 38-40. (Cai Xuanxuan. SECI and Knowledge-Based Streaming Living Library Activities[J]. Sci-tech Innovation and Productivity, 2013(8): 38-40.)

[39] 冯锐, 王克本. 高校图书馆的业务流程外包和知识流程外包[J]. 中华医学图书情报杂志, 2012, 21(4): 36-39. (Feng Rui, Wang Keben. Professional Work Outsourcing and Knowledge Process Outsourcing in Academic Libraries[J]. Chinese Journal of Medical Library and Information Science, 2012, 21(4): 36-39.)

[40] 江亮.基于知识管理的图书馆知识流程研究[J].情报探索, 2009(4): 26-28. (Jiang Liang. Research on KM-based Library Knowledge Process[J]. Information Research, 2009(4): 26-28.)

[41] 王月平. 战略联盟的知识流动循环模型研究[J]. 科技管理研究, 2010, 30(2): 157-159. (Wang Yueping. The Study of the Knowledge Flow Cycle Model of the Strategic Alliance[J]. Science and Technology Management Research, 2010, 30(2): 157-159.)

[42] 赵顺龙, 朱紫. 企业纵向合作中核心知识保护的路径与策略 [J]. 科技进步与对策, 2011, 28(24): 84-87. (Zhao Shunlong, Zhu Zi. Path and Strategies to Protect Core Knowledge in Enterprise Vertical Cooperation [J]. Science & Technology Progress and Policy, 2011, 28(24): 84-87.)

[43] 阎海峰. 知识在网络组织中的流动分析——以跨国网络结构为例[J]. 外国经济与管理, 2001, 23(10): 29-33. (Yang Haifeng. An Analysis of Knowledge Flow in Network Organization—Transnational Network Structure as an Example [J]. Foreign Economies & Management, 2001, 23(10): 29-33.)

[44] 丁立群, 李永周. 知识流动促进产业集群创新评价研究[J]. 工业技术经济, 2009, 28(11): 55-59. (Ding Liqun, Li Yongzhou. Evaluation of Innovation in Knowledge Flow to Promote Dustrial Clusters [J]. Industrial Technology & Economy, 2009, 28(11): 55-59.)

[45] 王生发, 顾新建, 张太华, 等. 协同设计中的知识流控制技术[J]. 农业机械学报, 2008, 39(11): 122-126. (Wang Shengfa, Gu Xinjian, Zhang Taihua, et al. Control Technology of Knowledge-flow in Collaborative Design [J]. Transactions of the Chinese Society for Agricultural Machinery, 2008, 39(11): 122-126.)

[46] 张晓刚, 李明树. 基于工作流的知识流建模与控制[J]. 软件学报, 2005, 16(2): 184-193. (Zhang Xiaogang, Li Mingshu. Workflow-Based Knowledge Flow Modeling and Control [J]. Journal of Software, 2005, 16(2): 184-193.)

[47] 朱卫未, 黄阳, 于娱, 等. 基于着色 Petri 网的组织知识流建模研究[J]. 计算机技术与发展, 2013, 23(4): 189-194. (Zhu Weiwei, Huang Yang, Yu yu, et al. Research on

Organizational Knowledge Flow Modeling Based on Colored Petri Net [J]. Computer Technology and Development, 2013, 23(4): 189-194.)

[48] 丁漪杰, 李孝忠. 基于着色Petri网的知识流建模与分析[J]. 系统仿真学报, 2011, 23(S1): 287-290. (Ding Yijie, Li Xiaozhong. Research on Knowledge Flow Modeling and Analysis Based on Colored Petri Net [J]. Journal of System Simulation, 2011, 23(S1): 287-290.)

[49] 赵晨, 陈国权, 高中华. 领导个人学习对组织学习成效的影响: 基于情境型双元平衡的视角[J]. 管理科学学报, 2014, 17(10): 38-49. (Zhao Chen, Chen Guoquan, Gao Zhonghua. The Effect of Team Leader Learning on Organizational Learning: A View Based on Contextual Ambidexterity[J]. Journal of Management Sciences in China, 2014, 17(10): 38-49.)

[50] 张成考, 吴价宝, 纪延光. 虚拟企业中知识流动与组织间学习的研究[J]. 中国管理科学, 2006, 14(2): 129-135. (Zhang Chengkao, Wu Jiabao, Ji Yan'guang. Study on Knowledge Flowing and Interorganizational Learning in Virtual Enterprise [J]. Chinese Journal of Management Science, 2006, 14(2): 129-135.)

[51] 李正风, 曾国屏. 创新系统理论中知识流分析的两个视角[J]. 科学学与科学技术管理, 2002, 23(4): 21-24. (Li Zhengfeng, Zeng Guoping. Two Angles of Theoretical Analysis in Innovation System Approaches[J]. Science of Science and Management of S.&.T., 2002, 23(4): 21-24.)

[52] 张瑞军, 李金玲. 区域创新系统中知识流动对经济增长的影响[J]. 辽宁石油化工大学学报, 2007, 27(4): 84-86. (Zhang Ruijun, Li Jinling. Investigations of Impact of Knowledge Flows in Regional Innovation System on Economic Performance [J]. Journal of Liaoning University of Petroleum & Chemical Technology, 2007, 27(4): 84-86.)

[53] 魏奇锋, 童洁, 顾新. 长三角区域创新体系的知识流动研究[J]. 经济问题探索, 2011(12): 35-40. (Wei Qifeng, Tong Jie, Gu Xin. Study on Knowledge Flow of Innovation System in the Yangtze River Delta Region [J]. Inquiry into Economic Issues, 2011(12): 35-40.)

[54] 黄西川, 韩玉启. 区域知识流动与技术创新——OECD 国家创新体系概念基本假设的实证分析[J]. 科学学与科学技术管理, 2003, 24(3): 24-27. (Huang Xichuan, Han Yuqi. Regional Knowledge Flow and Technical Innovation—OECD Countries Innovation System Concept [J]. Science of Science and Management of S.&.T., 2003, 24(3): 24-27.)

[55] 于腾, 王忠群. 面向认知协作的知识管理研究[J]. 计算机技术与发展, 2008, 18(4): 52-54. (Yu Teng, Wang Zhongqun. Cognitive-Cooperation-Oriented Knowledge Management Research [J]. Computer Technology and Development, 2008, 18(4): 52-54.)

[56] 窦万春, 刘茜萍, 蔡士杰. 面向认知协作的知识流分析与研究[J]. 计算机研究与发展, 2006, 43(6): 1109-1114. (Dong Wanchun, Liu Qianping, Cai Shijie. Cognitive-Cooperation-Oriented Knowledge Flow Research[J]. Journal of Computer Research and Development, 2006, 43(6): 1109-1114.)

[57] 周密, 承文, 韩立岩, 等. 知识流模型及其在航天企业中的应用[J]. 中国管理科学, 2005, 13(5): 81-88. (Zhou Mi, Cheng Wen, Han Liyan, et al. Practice of Applying Knowledge Flow Model in the Astronautic Enterprise [J]. Chinese Journal of Management Science, 2005, 13(5): 81-88.)

[58] 周密, 韩立岩. 知识流的 Petri 网模型[J]. 计算机工程与设计, 2005, 26(8): 2149-2152. (Zhou Mi, Han Liyan. Modeling Knowledge Flow with Petri Net[J]. Computer Engineering and Design, 2005, 26(8): 2149-2152.)

[59] 刘小军, 蒙大斌. 服务创新体系的知识流动及其政策建议[J]. 网络财富, 2009(6): 44-46. (Liu Xiaojun, Meng Dabin. Knowledge Flow and Its Strategies for Server Innovation System[J]. Internet Fortune, 2009(6): 44-46.)

[60] 章依凌, 陈晓英, 虞紫英. 嘉兴促进外贸生产企业转型升级研究[J]. 山东纺织经济, 2012(12): 69-70. (Zhang Yiling, Chen Xiaoying, Yu Ziying.Transformation and Upgrading of the Promotion of Foreign Trade Enterprises in Jiaxing[J]. Shandong Textile Economy, 2012(12): 69-70.)

[61] 王伟军, 刘艳芬. 知识流程外包产业的知识产权授权模式[J]. 情报理论与实践, 2011, 34(9): 6-9. (Wang Weijun, Liu Yanfen. Intellectual Property Licensing Mode for Knowledge Process Outsourcing Industry[J]. Information Studies: Theory & Application, 2011, 34(9): 6-9.)

[62] 朱胜勇. 印度服务外包发展现状及我国的比较与借鉴[J]. 国际经贸探索, 2008, 24(6): 71-76. (Zhu Shengyong. The Existing State of Indian Service Outsourcing and a Comparison with China [J]. International Economics and Trade Research, 2008, 24(6): 71-76.)

作者贡献声明:

王曰芬: 提出研究思路, 设计研究方案, 论文最终版本修订;
傅柱: 采集、整理、加工和分析数据, 起草论文;
陈必坤: 协助数据处理和分析, 参与修改论文。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: fuzhu886@163.com。

[1] 王曰芬, 傅柱, 陈必坤. OriginalData.txt. 国内知识流相关文献题录数据.

[2] 王曰芬, 傅柱, 陈必坤. ExperimentData.txt. 实验数据.

[3] 王曰芬, 傅柱, 陈必坤. F2/G3/F0/G4{G6}{G7}/G2/F4/TP3/C93/F1/TP1/F7/全局文件夹. 全局和 11 个分学科的 LDA 抽取中间数据.

[4] 王曰芬, 傅柱, 陈必坤. 学科分类和主题抽取结果数据.xlsx. 学科前期分类和主题抽取结果数据展示.

收稿日期: 2015-11-16
收修改稿日期: 2015-12-14

Analyzing Knowledge Structure Research with LDA Model

Wang Yuefen　Fu Zhu　Chen Bikun
(School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: [**Objective**] This paper aims to comprehensively explore the knowledge structure and hotspot distribution of different disciplines, with the help of topic extraction and distribution analysis using LDA (Latent Dirichlet Allocation) model from the perspective of subject classification. [**Methods**] We collected data from the domestic knowledge flow (KF) field and KF related literature from CNKI and Wangfang database, then grouped these data into 11 categories by Chinese Library Classification. Finally, we extracted 20 hot subjects from documents in 11 disciplines with the help of the LDA topic model. [**Results**] The content and knowledge points in 11 disciplines were obtained from the analysis of 20 extracted hot topics. [**Limitations**] We did not compare the proposed method with topic mining research in other fields. The domestic KF hotspots found by our study were not compared to the previous findings. [**Conclusions**] The proposed method can help us explore the knowledge structure and research trends more comprehensively.

Keywords: Knowledge flow　LDA　Topic extraction　Discipline classification　Structure map　Research focus